Final Report for AOARD Grant 114035 "Nano-Informatics"

Name of the Principal Investigators: Dr U.S.N.Murty
Head, Biology Division
CSIR-Indian Institute of Chemical Technology
Tarnaka, Hyderabad-607, INDIA
Email address: usnmurty@iict.res.in
Tel: +91-40-27160123-40
Period of performance: 2011-12 (sep 12)

ABSTRACT:

Nanoparticles (NP) are been used in our daily life, these helps humans in many ways, usage of these NP's had been increased tremendously these days without knowing its complications and health problems it could lead to, these nanoparticles should be studied for its toxicity, as in the case of nano-size property of every element is found to varying from its element in micro size, many nanoparticles was found to be highly reactive in its nano-size when compared to micro-sized element, so study on these nano sized elements is very much needed now for continued usage of its in medical field for imaging, delivery of drugs and as sunscreen (cosmetic) and etc

Development of prediction tool for nanotoxicity studies is an novel approach using the help of In-Silico techniques by which we can predict the nature of particular nanoparticle easily when compared to the traditional methods like In-Vivo in which animals are used as test organisms and In-vitro techniques in which cell lines are you used, Using animals for testing for In-vivo techniques, Ethical issues are raised, as animals are used as test organisms and many guidelines are to be followed for using animals for toxicity studies, Organization for Economic Co-operation and Development (OECD) has set guidelines for toxicity studies in guideline number 420, which say only dosage of 50-2000 mg/kg body weight can be used for the study of toxicity. In-vitro studies of toxicity by using cell lines is also very costly process, time taking process and laborious work requiring skillful people for carrying out these studies, as maintaining the cell lines properly is an big challenge and completing the toxicity studies on them is very hard.

By using the In–Silico tools for the predication we save time, money and even animals from sacrificing them, by which no ethical issues will be raised, and it's an easy process comparatively to in-vivo and in-vitro methods

OBJECTIVES: The main aim is to develop novel software on both cluster and classification that address the following specific objectives

1. Classification of nanomaterials (NM) using expert computer datamining techniques based on unique physicochemical properties of NM.
2. Clustering the NM by using SOM (Self Organizing Maps) and Weka depending on their size, shape, charge, and coating and other unique properties.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
16 NOV 2012	**Final**	**14-09-2011 to 13-09-2012**

4. TITLE AND SUBTITLE **Nanobioinformatics: emerging computational tools to understand nano-bio interaction**	5a. CONTRACT NUMBER **FA23861114035**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **USN Murty**	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Indian Institute of Chemical Technology,Tarnaka,Hyderabad, India,NA,NA,500007**	8. PERFORMING ORGANIZATION REPORT NUMBER **N/A**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **AOARD, UNIT 45002, APO, AP, 96338-5002**	10. SPONSOR/MONITOR'S ACRONYM(S) **AOARD**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) **AOARD-114035**

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Nanoparticles (NP) are being used in our daily life; these help humans in many ways; usage of these NP?s have increased tremendously recently without knowing their complications or health problems they could lead to; these nanoparticles should be studied for their toxicities, because the nano-size properties of every element are found to vary from the element in micro size, and many nanoparticles were found to be highly reactive in their nano-size when compared to the micro-sized element. Study of these nano sized elements is needed now for their continued usage in medical fields for imaging, delivery of drugs, as sunscreens (cosmetic) etc. Development of prediction tools for nanotoxicity studies is a novel approach using the help of in-silico techniques by which one can predict the nature of a particular nanoparticle easily when compared to traditional methods like in-vivo in which animals are used as test organisms and in-vitro techniques in which cell lines are used. In using animals for testing for in-vivo techniques, ethical issues are raised, as animals are used as test organisms and many guidelines need to be followed. E.g., in using animals for toxicity studies, the Organization for Economic Co-operation and Development (OECD) has set guidelines for toxicity studies in guideline number 420, which says that only dosages of 50-2000 mg/kg body weight can be used for the study of toxicity. In-vitro studies of toxicity by using cell lines is also a very costly and time-taking process, and laborious work requiring skillful people for carrying out these studies, such as maintaining the cell lines properly, is a big challenge, and completing the toxicity studies on them is very difficult to accomplish. By using the in?silico tools for the predication one saves time, money, and even animals from sacrificing them, by which no ethical issues will be raised, and it?s an easy process comparatively to in-vivo and in-vitro methods.

15. SUBJECT TERMS **Nanobio Devices**					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**5**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

3. Development of Structure-Toxicity Relationship Studies (STRS) using expert computer data mining techniques.
4. Development of novel tool interactions between nanoparticles of interest and biological-active compounds or proteins.
5. Final Report and Software submission.

INTRODUCTION:

Development of prediction tool for toxicity of nanoparticles requires huge amount of data, data based on which the tool is developed it can be generated by conducting the studies on the Wister male/female mice, these mice are very good test subject for carrying out the studies because these have similar genome as of the humans almost of 90%, these are easy to handle, morphological changes occurring in them can be easily noticed through naked eye.

Phase I: Data collection

Collected data on nanoparticles from the different manuscripts published in different journals, these articles are downloaded and data is collected from it. There are many online journals which report studies of NANOPATICLES toxicity these are: Nature Nanotechnology, SCIENCE DIRECT, ACS NANO, BMC, Small, Nano Today, Current Nano science, Nano Letters, Nanotechnology. The key words which were used for searching in the site are Nanotoxicity, Nano-particles, Nanoparticles AND toxicity. These where among the few words used for searching the articles. Huge number of articles have been found, all the relevant articles was downloaded and studied for the data. These articles where sorted based on the source and the particles upon which it is studied and sorted according to element and its source from where it is found like science direct, acs nano, wiley publications, etc.

Elements like Gold, silver, titanium are few among upon which most of the studies are concentrated, other elements are also reported but few numbers of articles had been found comparing to these elements, the elements in conjugated form had also been studied such as silver nitrate, titanium dioxide, and many more. Many studies had been reported until now as these NP individually are not potent but in conjugated form proved to be very dangerous to use. These data reported in the journals are collected in an excel sheet, each article deals with different nanoparticles and as each investigator will one conducts different experiments for his own purpose, this make challenging for the data collection and as every journal has different type of data represented in different fashion like tables, graphs, percentages, ratios and etc.

Phase –II About SOM tool

Self Organizing Maps (SOM)

Briefly, SOM is a data clustering technique invented by Professor Teuvo Kohonen of Helsinki University of Technology, Finland, in 1960's which reduce the dimensions of data through the use of self-organizing neural networks. In SOM the neurons are organized in a

lattice, usually a one or two-dimensional array, which is placed in the input space and is spanned over the inputs distribution. The processing units in the SOM lattice are associated with weights of the same dimension of the input data. Using the weights of each processing unit as a set of coordinates the lattice can be positioned in the input space. During the learning stage the weights of the units change their position and "move" towards the input points. This "movement" becomes slower and at the end of the learning stage the network is "frozen" in the input space. After the learning stage the inputs can be associated to the nearest network unit. When the map is visualized the inputs can be associated to each cell on the map. One or more cell that clearly contains similar objects can be considered as a cluster on the map. These clusters are generated during the learning phase without any other information. Hence, the main applications of the SOM is to visualize high-dimensional data in a two dimensional manner, and the creation of abstractions like in many clustering techniques.

The characteristic that distinguishes the SOM net from the other cluster algorithms is that not only similar inputs are associated to the same cell but also neighborhood cells contain similar types of documents. This property together with the easy visualization makes the SOM map a useful tool for visualization and clustering of large amount of data sets.

Steps involved in the algorithm

1. **Initialization:** Randomly initialize a weight vector (Wi) for each neuron i
   Wi = [wi1, wi2, ..., win], n denotes dimension of input data

2. **Sampling:** Select an input vector X = [x1, x2, ..., xn]
3.
4. **Similarity matching:** Find the winning neuron whose weight vector best matches with the input vector
   $j(t) = \arg\min \{ || X – Wi|| \}$

5. **Updating:** Update weight vector of winning neuron, such that it becomes still closer to the input vector. Also update weight vectors of neighbouring neurons – farther the neighbour, lesser the change.

   $Wi(t+1) = Wi(t) + \alpha(t) * hij(t) * [X(t) – Wi(t)]$
   $\alpha(t)$: Learning rate that decreases with time t, $0 < \alpha(t) <= 1$
   $hij(t) = \exp(-||rj – ri||2 / 2 * \sigma(t)2)$
   $||rj – ri||2$ = Distance between winning neuron and other neurons
   $\sigma(t)$ = Neighbourhood radius that decreases with time t

6. **Continuation:** Repeat steps 2 to 4 until there is no change in weight vectors or up to certain number of iterations For each input vector, find the best matching weight vector and allot the input vector to the corresponding neuron/cluster

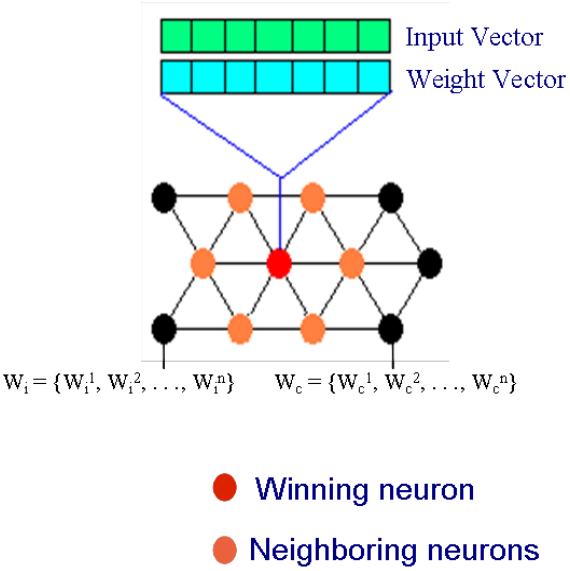

Illustration of an SOM neural network

Data Normalization:

Summarized data is normalized linearly in such a way that minimum value in each category is 0 and the maximum is 1. This is done to ensure that all the parameters are given to equal importance when clustering is done. The neuron weightage was adjusted by the learning rate. The learning rates and distance threshold values for the SOM are generally default values. Unsupervised learning was done using the data learning constant of 0.01 with 5000 iterations that yielded clusters based on the neighborhood distance.

Parameters identified for application of SOM:

Parameters are Size, Shape, Concentration, Exposure time, Mode of exposure, Surface group, present on it Stress response, Cell viability, Up regulation and down regulation of enzymes (GSH, SOD, GSSH, MDA, ALK, ALT, LDH), Cell lines.

Preprocessing: After collection of data from the published articles preprocessing of the data is done the process of pre processing include

- decide which fields to include in that data collected (document the rational for inclusion/exclusion)
- collect additional data if needed
- select data subsets to use
- consider use of sampling techniques to reduce size of data set decide if data balancing is required

SOM result:

Using SOM we can generate an image which is easy to study below is an example of possible result which we can get.

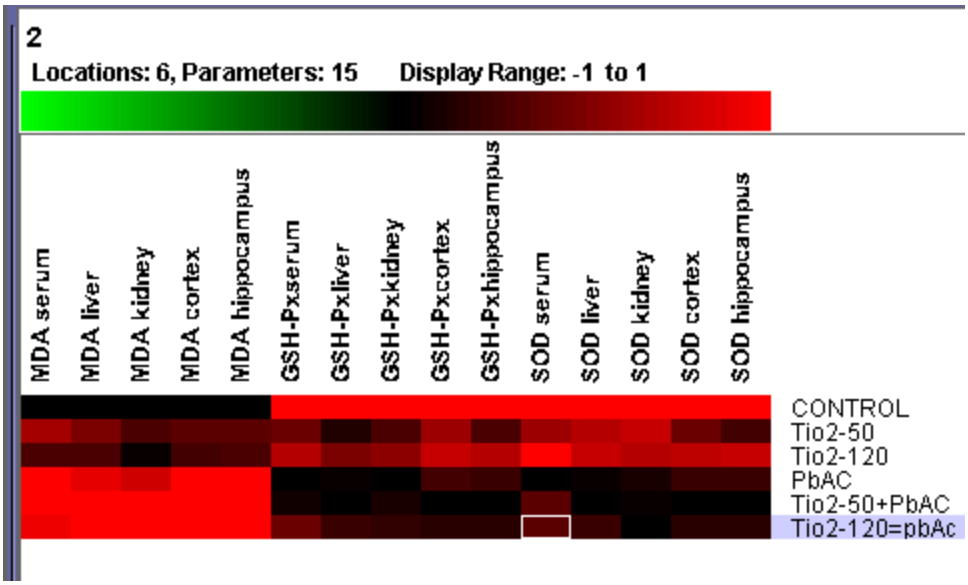

Figure-1: Result from SOM, shows that the stress response of Titaninum Dioxide when injected orally single dose with and without lead acetate, titanium dioxide when injected alone not much change in enzymes (Malondialdehyde, Glutathione reductase, superoxide Dismutase) concentration is seen (light region), but with conjugation of lead acetate on surface of TiO2 an elevated concentration of these enzymes found (dark regions).

Our goal is to develop an tool which helps in understanding NP nature, which can help in deciding the right amount of concentration for each purpose without causing the damage, we all know NP is used in imaging purpose in medical field, tool can help in deciding which NP will does the job without showing any harmful effect to the patient during the diagnosis, tool will help in deciding right concentration of dose to inject into a patient without causing toxicity. The SOM tool can employ for this purpose all of these can be used and have potential for used to develop the tool.

Future Plan: As the collection of data from the online available journals is going on, once the collection of data from those articles is completed, we can pre-process the data using methods described above; when the pre-processing of the data is finished we can move to phase-II

⇒ **Phase II:** Developing best classification and clustering methodologies:

- **Phase IIa**: Once the data collection is over they will be divided into training and test set. The whole data set will be normalized employing available standard statistical procedure before proceeding for the next step.
- **Phase IIb:** Application of Classification and clustering methodology